2006

Modeling the Firm-Size Distribution Using Box-Cox Heteroscedastic Regression

Zhenlin YANG
Singapore Management University, zlyang@smu.edu.sg

Yiu Kuen TSE
Singapore Management University, yktse@smu.edu.sg

DOI: https://doi.org/10.1002/jae.870

JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **21**: 641–653 (2006)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.870

MODELLING FIRM-SIZE DISTRIBUTION USING BOX–COX HETEROSCEDASTIC REGRESSION

Z. L. YANG AND Y. K. TSE*
School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road, Singapore 178903

SUMMARY

Using the Box–Cox regression model with heteroscedasticity (BCHR), we re-examine the size distribution of the Portuguese manufacturing firms studied by Machado and Mata (2000) using the Box–Cox quantile regression (BCQR) method. We show that the BCHR model compares favourably against the BCQR method. In particular, the BCHR model can answer the key questions addressed by the BCQR method, with the advantage that the estimated quantile functions are monotonic. Furthermore, estimation of the BCHR model is straightforward and the confidence intervals of the BCHR regression quantiles are easy to compute.

1. INTRODUCTION

In a recent article Machado and Mata (MM hereafter) (2000) analysed the size distribution of manufacturing firms in Portugal. Using the Box–Cox quantile regression (BCQR) method they examined and tested some implications of Gibrat's Law, including the prediction that firm sizes are log-normally distributed. They argued that the usual linear regression model is unable to give a complete picture of the conditional size distribution, and that the shifts in the conditional location and scale are not sufficiently captured by incorporating heteroscedasticity into the model. Also, as the marginal effects of the covariates on the shape of the size distribution may vary at different points of the distribution, regression analysis is not flexible enough to allow for these diverse effects. MM used the BCQR method to estimate the impact of the covariates on many attributes of the conditional distribution, such as the scale, skewness and kurtosis. They also proposed the use of marginal effects as a measure of the impact of the changes in the covariates on the quantiles, and maintained that classical linear regression models are unable to provide similar measures.

The purpose of this paper is to re-examine the use of the heteroscedastic regression model in analysing the firm-size distribution. Instead of imposing the logarithmic transformation, however, we adopt the flexible Box–Cox transformation. Thus, we consider the Box–Cox heteroscedastic regression (BCHR) model. We show that the BCHR model fits the firm-size data well and is able to answer the key questions addressed by MM. Furthermore, while the empirical conditional quantile function estimated by MM is not monotonic, this drawback does not occur in the BCHR method. Also, the estimated marginal effects of the covariates on the quantiles using the BCHR method are smooth functions, whereas estimates using the BCQR method are irregular and 'disparate'. Thus, we are able to draw coherent inferences about the marginal effects of the covariates on firm size, which was lacking in MM's study. Finally, the BCHR method is computationally easier than the

* Correspondence to: Y. K. Tse, School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road, Singapore 178903. E-mail: yktse@smu.edu.sg
Contract/grant sponsor: Wharton-SMU Research Center, Singapore Management University.

Received 21 March 2004
Revised 7 March 2005

BCQR method. Inference concerning the quantiles is not restricted to the choice set of quantiles in the estimation stage.

The rest of the paper is organized as follows. Section 2 summarizes the BCQR method and MM's results on analysing the Portuguese firm-size distribution. We also outline the BCHR estimation method, as well as the calculation of the confidence intervals of the quantiles and the marginal effects of the covariates. The data and the empirical results are described in Section 3. Section 4 concludes and discusses possible applications of the BCHR model and extensions.

2. BCQR AND BCHR ANALYSIS OF FIRM SIZE

Let y denote the response variable of interest and x a vector of k covariates. The p-quantile of y at x is defined by $y_p(x) = \inf\{y|F(y|x) \geq p\}$, where $F(y|x)$ is the cumulative distribution function of y given x. Consider a sample of data $\{y_i, x_i\}$ for $i = 1, \ldots, n$, the quantile regression model with transformation specifies that

$$h(y_p(x_i), \lambda_p) = x_i'\beta_p \tag{1}$$

where $h(\cdot, \lambda_p)$ is a monotonic one-to-one transformation that depends on a parameter λ_p, and β_p is a vector of regression coefficients. Note that in this model both the regression parameter β_p and the transformation parameter λ_p vary with p. A common example of the transformation $h(\cdot)$ is the well-known Box–Cox transformation. If we let $g(\cdot, \lambda_p)$ be the inverse of $h(\cdot, \lambda_p)$, then $y_p(x_i) = g(x_i'\beta_p, \lambda_p)$. Buchinsky (1998) provides a survey of the techniques and applications of the quantile regression model. Chamberlain (1994) and Buchinsky (1995) discuss the method of estimation for the parameters β_p and λ_p.

Given a covariate vector x, the marginal effect on the p-quantile under the BCQR model is defined as

$$m_p(x) = \frac{\partial y_p(x)}{\partial x} = \left[\frac{\partial g(t, \lambda_p)}{\partial t}\right]\beta_p \tag{2}$$

Note that $m_p(x)$ is a k-element vector, the jth element of which is the change in the p-quantile per unit increase in the jth covariate. It describes the marginal effects of the covariates at different quantile points of the size distribution and is one of the main aims of modelling firm-size distribution using the BCQR approach proposed by MM.

MM's data set consists of manufacturing firms in Portugal in 1983 and 1991. There are 18 552 firms in 1983 and 26 515 firms in 1991. These firms were operating in 155 industries. The response variable is Firm Size (as measured by employment) and the covariates are: Age, Growth, Patents, Imports, Exports, MES (minimum efficient scale), Turbulence, Industry Size and State.[1] MM focused on the conditional distribution and quantile function evaluated at the sample mean of the covariates.[2] They found that the firm-size distribution in 1991 has shifted to the left of 1983. Their

[1] Age (of the firm) affects size through Gibrat's Law. MES and Industry Size reflect economies of scale. Patents measures product differentiation. Imports and Exports capture the international trade environment. Growth (of industry employment) and Turbulence (a measure of entry and exit rates) measure the industry's dynamics. State is a proxy for state intervention. See Machado and Mata (2000) for the details. Except for Age, which is firm-specific, all other covariates are industry-specific.

[2] The term 'quantile function' is used to refer to $y_p(\bar{x})$, i.e., the conditional quantile evaluated at the mean of the covariates. This notion of *conditional* quantile will be adopted in subsequent discussions.

estimated quantile functions, however, are *not monotonic* in p.[3] Also, the marginal-effect functions as evaluated by equation (2) are irregular and not smooth in p. Furthermore, the characteristics of the conditional distribution (scale, skewness and kurtosis) computed from the quantile function are 'rather unstable' and 'quite different from 1983 to 1991'.

MM's analysis of the conditional distribution depends on the estimated quantiles using the BCQR method. We shall show that the quantiles can also be estimated using the BCHR method. As a consequence, we can compute other characteristics of the conditional distribution, as well as the marginal effects of the covariates, as proposed by MM. Apart from the simplicity in the estimation of the BCHR model as compared to the BCQR model, which requires two-step iterations over the transformation and regression parameters, the estimation of the marginal effects and the effects of the covariates on other characteristics for *any* p-quantile under the BCHR model are straightforward once the estimated model is obtained. In contrast, for the BCQR model, inference is only available for the p-quantiles that are included in the estimation stage.

The BCHR model specifies the following equation for the response variable:

$$h(y_i, \lambda) = x_i'(\lambda)\beta + \sigma\omega(v_i, \gamma)e_i, \quad i = 1, \ldots, n \tag{3}$$

where $x_i'(\lambda) = \{x_{1i}', x_{2i}'(\lambda)\}$ is the covariate vector. Thus, x_{1i} is the vector of covariates which are not transformed (these may include dummy variables and time trend) and $x_{2i}(\lambda)$ is the vector of covariates which are transformed (with the same transformation function as the response variable). The model allows for heteroscedasticity, with v_i being the vector of weighting variables, σ the scaling factor for the standard deviation of the errors, and ω the weighting function with parameter γ. Note that v_i may or may not overlap with x_i. We assume the e_i are i.i.d. $N(0, 1)$.

Let $y_p(x_i, v_i)$ be the p-quantile of y at x_i and v_i, and z_p be the p-quantile of e_i. As the e_i are i.i.d., z_p is a constant across i. As h is a one-to-one transformation, we have

$$h(y_p(x_i, v_i), \lambda)) = x_i'(\lambda)\beta + \sigma\omega(v_i, \gamma)z_p \tag{4}$$

which implies, after an inverse transformation, $y_p(x_i, v_i) = g(x_i'(\lambda)\beta + \sigma\omega(v_i, \gamma)z_p, \lambda)$. A point estimator for $y_p(x_i, v_i)$ can be obtained by plugging the maximum likelihood estimates (MLE), denoted by hats, for the parameters. Thus, we have

$$\hat{y}_p(x_i, v_i) = g(x_i'(\hat{\lambda})\hat{\beta} + \hat{\sigma}\omega(v_i, \hat{\gamma})z_p, \hat{\lambda}) \tag{5}$$

which is a consistent estimate of the conditional quantile. Note that $\hat{y}_p(x_i, v_i)$ is a monotonic function of p. Yang and Tse (2002) proposed a confidence interval for the conditional quantile, called the corrected plug-in quantile limits (CPQL). Their method, though developed for a model with no transformed exogenous variables, can easily be extended to the model considered in this paper. The formulae for calculating the MLE and the CPQL when some of the covariates are subject to the same transformation as the dependent variable are given in a longer version of this paper.[4] In particular, the CPQL can also be used to calculate a confidence interval for the quantile of an out-of-sample covariate vector x_0 and a weighting vector v_0, serving the prediction purpose.

[3] Bassett and Koenker (1982) showed that the conditional quantile functions of the linear model evaluated at the sample mean of the covariates are monotonic in p. This result, however, does not hold for the case when the response variable is transformed.
[4] The paper is available from www.sess.smu.edu.sg/research/research_introduction.htm. Yang and Tse (2002) provide the derivation of the CPQL for regression models with no transformation in the covariates. They present some Monte Carlo results to show that the CPQL performs very well in small samples compared to the delta method.

To measure the effects of the covariates on the firm-size distribution, we calculate the marginal effects of the covariates on the p-quantile, i.e., $m_p(x, v) = \partial y_p(x, v)/\partial x$, given by

$$m_p(x, v) = (1 + \lambda[x'(\lambda)\beta + \sigma\omega(v, \gamma)z_p])^{(1-\lambda)/\lambda}[(\partial x(\lambda)/\partial x) \odot \beta + \sigma z_p(\partial v'/\partial x)(\partial\omega(v, \gamma)/\partial v)] \quad (6)$$

where $\odot$ is the Hadamard product of two vectors (i.e., the elementwise multiplication operator). Finally, the scale, skewness and kurtosis of the conditional distribution can be calculated from the conditional quantiles (see Machado and Mata, 2000, table II). The marginal effects of the covariates on these characteristics can be calculated by differentiating these quantities numerically with respect to x.

3. EMPIRICAL RESULTS

We estimate the BCHR model for the two years of firm data separately. As the covariate Industry Size is of a similar nature as the dependent variable Firm Size, there may be a case for transforming this variable in the regression. We fitted models with and without transformation for Industry Size. The results for the two cases are similar in many aspects (parameter estimates as well as diagnostics). The model with no transformation on Industry Size, however, provides tighter intervals for the conditional quantiles. Also, the results reported by MM appear to be based on the case with no transformation on Industry Size. Thus, in what follows we present only the results with no transformation on the covariates.[5]

We first estimate a Box–Cox transformation model with homoscedastic errors and test the homoscedasticity assumption using the Lagrange Multiplier (LM) test with all covariates taken as the weighting variables. The LM statistic is asymptotically distributed as a χ^2 with 9 degrees of freedom (the number of weighting variables) under the null of homoscedasticity, which is convincingly rejected with the LM statistic being 270.8 for 1983 and 359.0 for 1991. Following Yu *et al*. (2003) we plot the conditional p-quantiles of the Firm Size against the covariates for different values of p. These quantiles are expected to be parallel if the errors are homoscedastic and transformations are not necessary (i.e., $\lambda = 1$). Figure 1 plots the conditional quantiles of the Box–Cox homoscedastic regression model for 1983 against Age and MES. The results suggest strongly that the distributions of firm sizes are heteroscedastic.

We proceed to estimate the Box–Cox regression model with heteroscedasticity. In particular, we assume multiplicative heteroscedasticity in which the weighting function is the exponential function, i.e., $\omega(v_i, \gamma)^2 = \exp(v_i'\gamma)$. We assume the weighting variables v_i consist of the whole set of covariates x_i. This assumption is not expected to affect the main conclusions of the study. As seen below, most of the covariates are indeed significant weighting variables.

Table I summarizes the results of the estimated BCHR models for 1983 and 1991. Almost all covariates are significantly different from zero at the 5% level (the only exception is State in 1983). All the weighting variables are statistically significant at the 5% level except for Growth and Imports.[6] The null hypothesis of a linear functional relationship (i.e., $\lambda = 1$) is strongly rejected

[5] Results for the model with Industry Size transformed can be obtained from the authors on request.
[6] This is true for both years. It is interesting to observe that Exports, but not Imports, is significant for heteroscedasticity. Thus, higher exports induce larger *variations* in firm size (not just larger firm size), probably through providing scope and opportunity for growth. In contrast, the effects of imports on firm size appear to be of first order (affects firm size negatively) and not second order (no effects on *variation* of firm size).

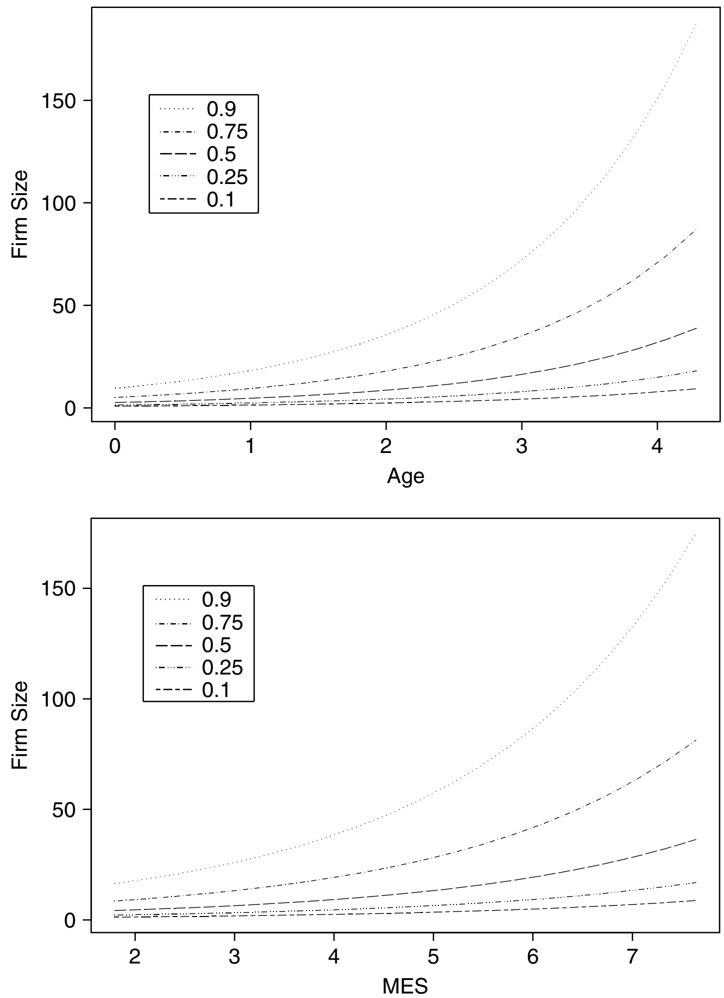

Figure 1. Conditional quantile versus Age and MES (Box–Cox homoscedastic regression, 1983)

with the LM statistic being 5485.66 for 1983 and 3048.77 for 1991. Also, the null hypothesis of a log-linear functional relationship (i.e., $\lambda = 0$) is soundly rejected with the LM statistic being 139.75 and 277.57 for 1983 and 1991, respectively. The regression coefficients of the covariates over the two years are quite similar, except for Turbulence and State.

Figure 2 plots the empirical density functions of the standardized residuals of the BCHR model against the standard normal density for comparison. The empirical density functions of the standardized residuals over the two years overlap a lot, and are very close to the standard normal density. Figure 3 plots the quantile functions evaluated at the sample means of the covariates, i.e., $\hat{y}_p(\overline{x}, \overline{v})$. Apart from the conditional quantile functions for the 1983 and 1991 data, we also plot the conditional quantile function based on the estimated 1983 model evaluated at the sample mean of the 1991 data, labelled as 1983–91. It is clear from the plots that the distribution of firm

Table I. Estimated BCHR models

	1983		1991	
	Estimate	Standard error	Estimate	Standard error
Covariates				
Intercept	−0.7420	0.0755	−0.5699	0.0673
Age	0.5448	0.0096	0.4980	0.0074
Growth	0.6401	0.0939	0.3477	0.0745
Patents	9.3959	1.8281	8.7646	1.6978
Imports	−0.0227	0.0033	−0.0269	0.0033
Exports	0.3142	0.0252	0.2942	0.0144
MES	0.3121	0.0114	0.2893	0.0109
Turbulence	−11.0892	1.8625	−61.8533	3.4545
Industry Size	0.0243	0.0070	0.0222	0.0059
State	0.0211	0.1348	0.3364	0.1361
Weighting variables				
Age	0.1046	0.0064	0.0605	0.0050
Growth	−0.0387	0.0614	0.0524	0.0456
Patents	2.8741	1.2578	2.4820	1.2306
Imports	−0.0009	0.0024	0.0040	0.0024
Exports	0.0391	0.0164	0.0234	0.0112
MES	0.0753	0.0084	0.0947	0.0079
Turbulence	3.3004	1.3880	18.3768	1.8469
Industry Size	−0.0156	0.0054	−0.0352	0.0046
State	0.3944	0.0737	0.4682	0.0803
Transformation and scale parameters				
λ	−0.0655	0.0055	−0.0816	0.0047
σ	0.5735	0.0366	0.6618	0.0453

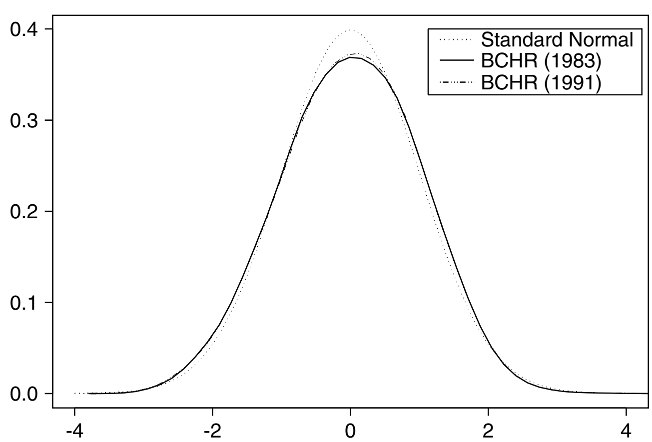

Figure 2. Density functions of standardized residuals of firm size

size has shifted to the left from 1983 to 1991. In other words, there were (conditionally) more smaller firms in 1991 than in 1983. It can be seen that, however, the graph for 1983–91 has shifted upwards versus the graph for 1983, which suggests that the difference between 1983 and 1991 is due to the change in the effects of the covariates on the firm size rather than the changes in the

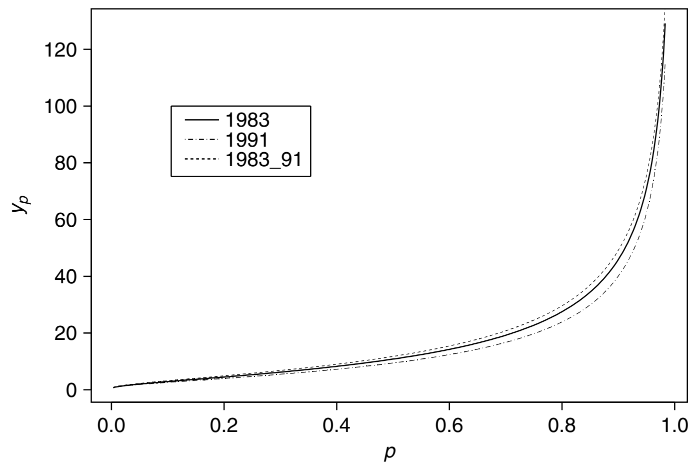

Figure 3. Conditional quantile functions

Table II. Descriptive statistics of the conditional distributions

		1983	1991	1983–91
Location	$y_{0.5}$	10.832	9.435	11.751
Scale	$(y_{0.5} - y_{0.25})/(y_{0.75} + y_{0.25})$	0.620	0.616	0.616
Skewness	$(y_{0.5} + y_{0.25} - 2y_{0.5})/(y_{0.75} - y_{0.25})$	0.369	0.371	0.366
Kurtosis	$(y_{0.90} - y_{0.10})/(y_{0.75} - y_{0.25})$	2.466	2.485	2.455

covariates. Figure 3 provides similar results to those of MM (see figure 3 of MM). However, one important drawback of MM's empirical conditional quantile is that it is *not monotonic*.

Table II summarizes some summary statistics of the conditional (on the sample mean of the covariates) distributions.[7] This can be compared against table II of the BCQR results of MM. Similar to MM, the location has shifted to the left, the scale has become smaller, and the tail has become thicker. In contrast to MM, however, the skewness has shifted slightly to the right. Thus, similar to the BCQR method, the BCHR method is able to provide additional information about the shape of the conditional distribution.

Table III provides the point estimates as well as the 95% confidence intervals of the conditional quantiles. We observe that the confidence intervals get wider for larger p. Also, the intervals of all percentiles over the two years do not overlap, showing that the conditional quantiles change significantly over time. In contrast, the confidence intervals of the conditional quantiles are not available in MM. Table IV presents the estimated marginal effects at selected values of p. The results can be compared against table III of MM. It can be seen that the marginal effects of all covariates are statistically significant at the 5% level except for Industry Size and State. While Industry Size is insignificant in the higher quantiles, State is significant in the higher quantiles. This finding shows that industry-wide economies of scale have little impact on the big firms, while state interventions are mostly restricting the big firms. In contrast, MM's results show 'considerable

[7] The set of values p chosen follow those of MM. This is to facilitate comparison with MM's results. This remark applies to other tables below.

Table III. Conditional quantile estimates and 95% confidence intervals

p (%)	1983			1991		
	Lower limit	Point estimate	Upper limit	Lower limit	Point estimate	Upper limit
1	1.032	1.068	1.104	0.971	0.991	1.011
10	2.839	2.896	2.951	2.551	2.592	2.632
20	4.417	4.497	4.576	3.921	3.978	4.033
30	6.117	6.227	6.335	5.397	5.469	5.541
40	8.122	8.269	8.417	7.135	7.228	7.321
50	10.637	10.832	11.031	9.294	9.435	9.579
60	13.997	14.258	14.528	12.225	12.390	12.558
70	18.878	19.239	19.617	16.460	16.697	16.942
80	26.990	27.529	28.102	23.521	23.899	24.296
90	44.875	45.901	47.007	39.256	40.014	40.822
95	69.026	70.948	73.043	60.910	62.292	63.779
99	158.954	166.318	174.502	145.147	149.586	154.426

Table IV. Marginal effects of covariates on conditional quantiles

p (%)	Age	Growth	Patents	Imports	Exports	MES	Turbulence	Industry size	State
					1983				
10	1.308	2.129	18.652	−0.067	0.832	0.693	−46.507	0.132	−1.378
	(0.038)	(0.405)	(6.942)	(0.013)	(0.081)	(0.045)	(8.817)	(0.030)	(0.500)
25	2.856	3.952	45.305	−0.132	1.725	1.579	−78.181	0.202	−1.330
	(0.060)	(0.646)	(11.403)	(0.021)	(0.140)	(0.073)	(13.565)	(0.048)	(0.849)
50	6.897	8.103	118.954	−0.287	3.997	3.951	−140.391	0.308	0.268
	(0.119)	(1.188)	(23.144)	(0.042)	(0.318)	(0.144)	(23.509)	(0.089)	(1.707)
75	17.034	17.212	312.339	−0.649	9.456	10.024	−252.614	0.410	7.428
	(0.314)	(2.620)	(57.317)	(0.104)	(0.859)	(0.349)	(50.015)	(0.207)	(3.974)
90	36.396	35.053	753.840	−1.398	21.246	23.636	−424.473	0.352	28.664
	(0.955)	(6.281)	(144.641)	(0.265)	(2.251)	(0.901)	(121.711)	(0.525)	(9.451)
95	66.933	54.534	1288.976	−2.250	35.016	39.952	−571.641	0.071	57.949
	(1.957)	(11.115)	(258.207)	(0.477)	(4.703)	(1.684)	(219.235)	(0.954)	(16.424)
					1991				
10	1.203	0.807	16.662	−0.088	0.750	0.509	−231.665	0174	−0.546
	(0.035)	(0.198)	(5.058)	(0.012)	(0.062)	(0.036)	(14.581)	(0.022)	(0.472)
25	2.463	1.687	38.831	−0.156	1.494	1.241	−388.328	0.231	0.303
	(0.049)	(0.362)	(8.600)	(0.019)	(0.090)	(0.061)	(22.369)	(0.036)	(0.785)
50	5.644	3.940	99.328	−0.305	3.335	3.279	−700.978	0.252	3.813
	(0.082)	(0.845)	(19.242)	(0.037)	(0.163)	(0.123)	(38.670)	(0.067)	(1.543)
75	13.465	9.553	258.302	−0.618	7.770	8.720	−1282.582	0.301	15.530
	(0.281)	(2.247)	(51.934)	(0.095)	(0.448)	(0.315)	(78.041)	(0.163)	(3.573)
90	30.643	22.017	626.267	−1.210	17.350	21.459	−2217.356	−0.956	46.924
	(1.040)	(5.759)	(137.138)	(0.252)	(1.350)	(0.904)	(173.044)	(0.444)	(8.691)
95	51.157	37.007	1080.507	−1.841	28.662	37.297	−3062.983	−2.531	88.884
	(2.244)	(10.311)	(250.047)	(0.461)	(2.661)	(1.821)	(295.260)	(0.855)	(15.523)

Note: Figures in parentheses are standard errors.

diversity', finding almost all possible cases, ranging from variables which are significant at all quantiles to those which are significant at a single quantile, and some which are significant only in the middle range.

Figure 4 plots the marginal effects of various covariates against p and can be compared against figure 4 of MM. Overall the shapes of our marginal-effect curves are similar to those of MM, except for Industry Size. However, while the BCHR method provides smooth marginal-effect curves, those computed using the BCQR method are erratic. This appears to arise from the instability of the algorithm rather than the nature of the underlying structure. Note that similar to MM, we find that the covariates Imports, Turbulence and Industry Size have quite different marginal-effect curves compared to others. First, the marginal effects of these covariates decrease as p increases. Second, the marginal effects of Turbulence are quite different over the two years. In contrast to MM, however, we find the effects of Industry Size over the two years to be similar and decreasing when p is large.

Following MM, we provide results of tests for the equality of marginal effects over adjacent quantiles as well as over the two different years. The results are summarized in Tables V and VI, which provide the t-ratios of the tests. Similar results using the BCQR method can be found in tables IV and V of MM. Table V shows that most t-ratios are significant at the 5% level, except for the covariate Industry Size. In contrast, many of the t-ratios in MM's results are statistically insignificant. In particular, none of the marginal effects between the 50th and 75th percentiles are statistically significant in 1991. In view of the differences in the results between 1983 and 1991, as well as the results in other quantiles, this finding appears to be anomalous. Indeed, MM commented that 'the majority of the covariates exert rather disparate effects across the distribution and that these have changed over time'. In contrast, we note that the BCHR results are quite stable across different quantiles. Furthermore, the computation of the asymptotic variance in the BCHR model is quite straightforward, while that of the BCQR method requires the delta method. It is not sure whether the apparent anomaly is due to the computational instability.

From Table VI we can see that the marginal effects of Age, Growth, MES and Turbulence are largely significantly different over the two years. This result is in agreement with that of MM. Our results, however, show coherency and consistency in signs, while this cannot be said of MM's BCQR results. For example, there are several irregular changes in signs of the t-ratios in table V of MM.

Finally, Table VII summarizes the marginal effects of the covariates on the scale, skewness and kurtosis of the conditional distribution, with standard errors in parentheses. While table VI of MM produces some results on the 'marginal' effects of the covariates on these attributes, they are computed using a different method (by taking a one-standard-deviation variation in the covariates). Thus, we will not compare our results against those of MM, except in concurring with MM that their 'estimated effects are rather unstable'. In contrast, our results show coherency and stability over the two years. In particular, Age, Patents, Exports, MES, Industry Size and State have significant marginal effects on the distribution attributes in both years, while Growth and Imports are insignificant. Turbulence is significant in 1991, but not in 1983.

4. CONCLUSIONS, OTHER APPLICATIONS AND EXTENSIONS

We have examined the use of the BCHR model for analysing firm-size distribution of manufacturing firms in Portugal. Following the work of Machado and Mata (2000), we estimate the conditional quantile function, the scale, skewness and kurtosis of the conditional distribution, the marginal effects of the covariates on the quantiles as well as on the scale, skewness and kurtosis. We have shown that the BCHR model provides a useful alternative to the BCQR model in

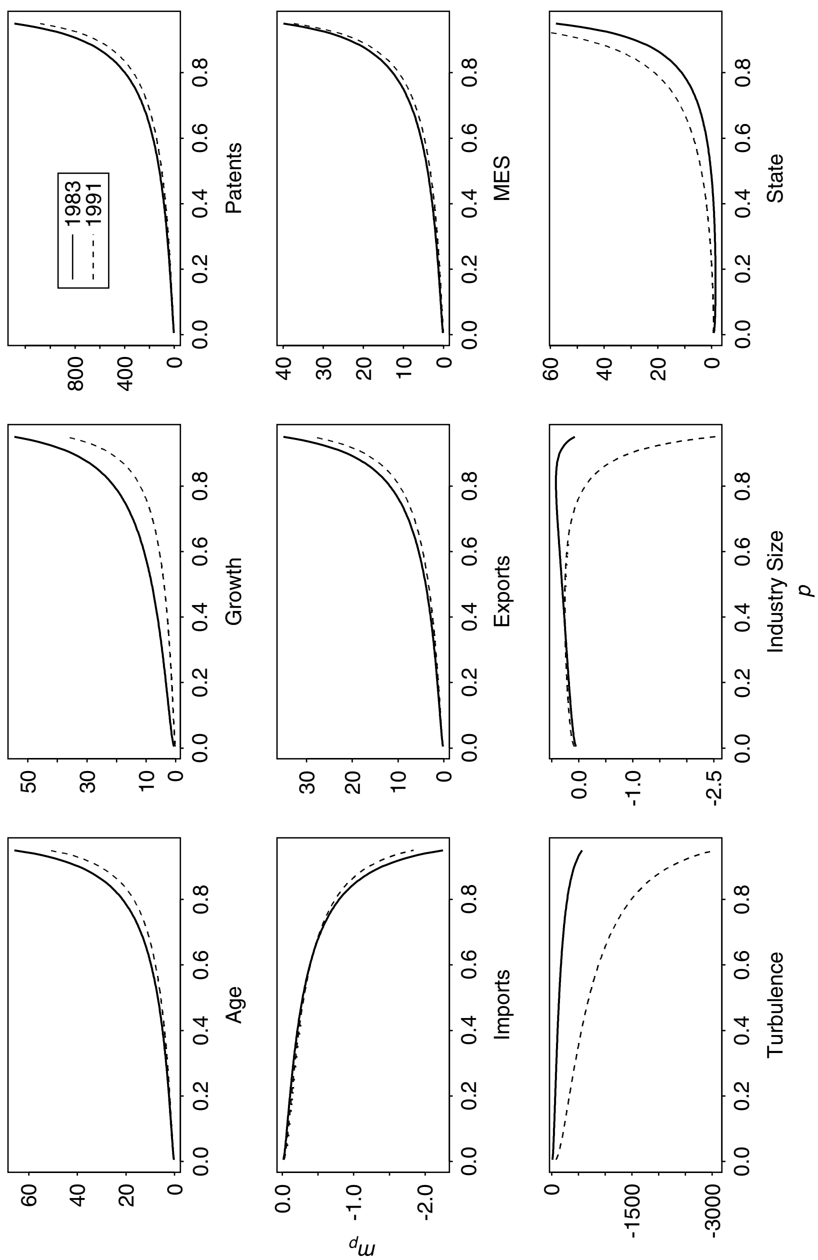

Figure 4. Plots of marginal effects

Table V. t-statistics for tests of equality of adjacent marginal effects

p (%)	Age	Growth	Patents	Imports	Exports	MES	Turbulence	Industry size	State
					1983				
10–25	55.871	6.813	5.126	−6.853	12.530	27.124	−5.973	3.483	0.124
25–50	56.124	6.655	5.454	−6.375	11.234	28.714	−5.215	2.167	1.694
50–75	45.477	5.674	5.242	−5.361	9.607	26.973	−3.610	0.763	2.957
75–90	32.726	4.585	4.885	−4.472	8.277	23.280	−2.217	−0.172	3.735
90–95	25.087	3.895	4.623	−3.944	7.454	19.997	−1.451	−0.636	4.100
					1991				
10–25	65.965	4.667	5.156	−8.157	20.438	26.200	−18.083	3.735	2.464
25–50	55.123	4.307	5.018	−6.687	18.261	28.537	−16.551	0.574	4.179
50–75	33.488	3.865	4.613	−4.944	13.536	25.721	−12.894	−2.055	5.362
75–90	21.955	3.494	4.226	−3.666	10.206	20.446	−9.048	−3.323	5.876
90–95	16.819	3.263	3.974	−2.965	8.482	16.697	−6.568	−3.726	5.967

Table VI. t-statistics for tests of equality of marginal effects between 1983 and 1991

p (%)	Age	Growth	Patents	Imports	Exports	MES	Turbulence	Industry size	State
10	2.032	2.933	0.232	1.187	0.804	3.193	10.866	−1.129	−1.210
25	5.073	3.059	0.453	0.847	1.388	3.553	11.855	−0.483	−1.412
50	8.670	2.856	0.652	0.322	1.797	3.548	12.387	0.503	−1.541
75	8.470	2.219	0.699	−0.220	1.740	2.774	11.112	1.438	−1.516
90	6.199	1.530	0.640	−0.514	1.484	1.706	8.475	1.902	−1.422
95	4.943	1.156	0.580	−0.617	1.306	1.070	6.774	2.031	−1.369

Table VII. Marginal effects on distribution attributes

	Age	Growth	Patents	Imports	Exports	MES	Turbulence	Industry size	State
					1983				
Scale	0.065	0.005	1.606	−0.001	0.028	0.044	1.096	−0.006	0.177
	(0.002)	(0.027)	(0.570)	(0.001)	(0.008)	(0.004)	(0.604)	(0.002)	(0.033)
Skewness	0.049	0.003	1.211	−0.001	0.021	0.033	0.826	−0.005	0.133
	(0.002)	(0.020)	(0.429)	(0.001)	(0.006)	(0.003)	(0.455)	(0.002)	(0.025)
Kurtosis	0.179	0.013	4.392	−0.003	0.077	0.121	2.997	−0.017	0.484
	(0.008)	(0.074)	(1.552)	(0.003)	(0.022)	(0.010)	(1.648)	(0.007)	(0.091)
					1991				
Scale	0.049	0.039	1.493	0.0001	0.023	0.055	5.502	−0.015	0.224
	(0.002)	(0.022)	(0.578)	(0.001)	(0.005)	(0.003)	(0.757)	(0.002)	(0.036)
Skewness	0.037	0.031	1.125	0.000	0.019	0.042	4.008	−0.010	0.170
	(0.001)	(0.017)	(0.437)	(0.001)	(0.004)	(0.002)	(0.564)	(0.002)	(0.027)
Kurtosis	0.139	0.110	4.264	0.002	0.067	0.157	15.709	−0.042	0.640
	(0.008)	(0.062)	(1.647)	(0.003)	(0.016)	(0.009)	(2.098)	(0.007)	(0.103)

Note: Figures in parentheses are standard errors.

analysing firm-size distribution. It has the advantage of producing monotonic empirical quantile functions. Furthermore, the results on the marginal effects of the covariates are more coherent and provide a more stable description of these effects compared to the BCQR method.

The BCHR model may be applied to other areas of research interest, including income distribution, health-care expenses and wage structure. Income inequality has recently developed into a major research area. Current studies in the literature adopt subclassification of the sample to examine within and between group inequality. This methodology, however, may not appropriately capture the interaction between groups. In contrast, the BCHR model provides a method to more efficiently extract information and formal hypothesis testing can be conducted under the BCHR model. Also, the income inequality literature has made use of scalar measures such as the Gini coefficient, variance of natural log of earnings and coefficient of variation. Some of these measures (such as the coefficient of variation) can be expressed or approximated by functions of the quantiles. Detailed analysis can then be conducted on the marginal impact of the covariates on these (approximate) scalar inequality measures using the BCHR model.

Health-care and wage data are well known to be skewed and non-normal (see Lambert and Larcker, 1995; Buchinsky, 1995), and may be analysed using the BCHR model. Tuckman *et al.* (1999) considered a regression model in which the response (surplus per discharge) variable is transformed using the modulus transformation. Their model can be extended by allowing for heteroscedasticity, with analysis of the impact of the covariates on the quantiles and other distributional characteristics.

The multiplicative heteroscedasticity model is used in this paper due to its popularity in the literature and its easy control of the non-negativity condition in estimation. For further discussion of the choice of the heteroscedasticity function, see Carroll and Rupert (1988). The BCHR model may also be extended beyond the Box–Cox transformation. In particular, as economic data are not always positive, a more general transformation that allows for both positive and negative observations may be required.

ACKNOWLEDGEMENTS

We are indebted to the Portuguese Ministry of Employment and Jose Mata for making the manufacturing firms' data available, and Chenwei Li for excellent research assistance. Research support from the Wharton-SMU Research Center, Singapore Management University, is gratefully acknowledged. We thank John Rust, the Editor, and three anonymous referees for their many helpful comments. The usual disclaimer applies.

REFERENCES

Bassett G Jr, Koenker R. 1982. An empirical quantile function for linear models with iid errors. *Journal of the American Statistical Association* **77**: 407–415.

Buchinsky M. 1995. Quantile regression, Box–Cox transformation model, and the U.S. wage structure, 1963–1987. *Journal of Econometrics* **65**: 109–154.

Buchinsky M. 1998. Recent advances in quantile regression models: a practical guideline for empirical research. *Journal of Human Resources* **33**: 88–126.

Carroll RJ, Ruppert D. 1988. *Transformation and Weighting in Regression*. Chapman and Hall: New York.

Chamberlain G. 1994. Quantile regression, censoring and the structure of wages. In *Advances in Econometrics*, Sims C (ed.). Cambridge University Press: New York; 171–209.

Lambert RA, Larcker DF. 1995. The prospective payment system, hospital efficiency, and compensation contracts for senior-level hospital administrators. *Journal of Accounting and Public Policy* **14**: 1–31.

Machado JAF, Mata J. 2000. Box–Cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics* **15**: 253–274.

Tuckman HP, Chang CF, Okunade AA. 1999. A transformation-both-side modulus power model: an application in health care. *Applied Economic Letters* **6**: 741–745.

Yang ZL, Tse YK. 2002. A corrected plug-in method for the quantile confidence interval of a transformed regression. SMU Economics & Statistics Working Paper Series, #22-2002 (www.sess.smu.edu.sg/research/research_introduction.htm).

Yu K, Liu Z, Stander J. 2003. Quantile regression; applications and current research areas. *The Statistician* **52**: 331–350.

